Exponential inequalities for dependent processes

Bernard Delyon

Exponential inequalities for dependent processes

Bernard Delyon
IRMAR, Université Rennes 1
Campus de Beaulieu, 35042 Rennes Cedex, France
bernard.delyon@univ-rennes1.fr

Abstract

We give exponential inequalities and Gaussian approximation results for sums of weakly dependent variables. These results lead to generalizations of Bernstein and Hoeffding inequalities, where an extra control term is added; this term contains conditional moments of the variables.

1 Introduction

In the whole paper $(X_i)_{1\leqslant i\leqslant n}$ is a sequence of *centred* random variables. Our objective is to give new exponential inequalities and Gaussian approximation results for the sum $S = X_1 + \cdots + X_n$ in the case where first and second order mixing conditions are assumed (first order mixing conditions involve conditional means and second order ones involve conditional covariances). This paper improves a previous one [9] by extending some results, simplifying the presentation, and correcting two errors: a term was forgotten in the expression of q Theorem 1, and another in the expression of w_p given in the remark following Theorem 9.

The essential application of exponential inequalities is to give small event probabilities; typically, we would like here to extend the Hoeffding and Bernstein inequalities to mixing processes, that is

$$P(S \geqslant A) \leqslant \exp\left(-\frac{2A^2}{\sum_i b_i^2 + \rho}\right), \qquad a_i \leqslant X_i \leqslant a_i + b_i \tag{1}$$

$$P(S \geqslant A) \leqslant \exp\left(-\frac{A^2}{2E[S^2] + 2Am/3 + \rho}\right), \qquad m = \sup_k \|X_k\|_\infty \tag{2}$$

where $\rho = 0$ if the variables are independent. We shall obtain here a value of ρ which depends on conditional moments of the variables. We also want to provide inequalities which generalize what is already known for martingales. Actually Equation (2) is not satisfied for a martingale ($E[S^2]$ has to be replaced with a bound on the total variation); this will lead us to two different alternatives

$$P(S \geqslant A) \leqslant \exp\left(-\frac{A^2}{2v + 2Am/3 + \rho}\right), \tag{3}$$

where v is a bound on some kind of quadratic variation and $\rho = 0$ in the case of a martingale (cf. Theorem 7 for a precise statement), or

$$P(S \geqslant A) \leqslant \exp\left(-\frac{A^2}{2E[S^2] + \sqrt{2Aw/3}}\right), \tag{4}$$

where w is a third order quantity (i.e. if S is replaced with tS, w becomes t^3w) which involves conditional moments and is in the independent case smaller than $\frac{3}{2}\sum\|X_k^3\|_\infty$ (Theorem 14). For instance, if we are close to the independent case and S is a normalized sum, that is $X_k = U_k/\sqrt{k}$ where $(U_k)_k$ is a sequence of weakly dependent bounded random variables, w has order $1/\sqrt{n}$ and the second term in the denominator is residual as long as $A \ll \sqrt{n}$.

Bounds like Equation (3) will be obtained through what we call here the first order approach whereas (4) will require the second order approach. We present the main ideas below.

First order approach for spatial processes. Considering the sequence $X_1, \dots X_n$ as a time series, for instance a martingale, it is natural to introduce the σ-fields

$$\mathcal{F}_k = \sigma(X_1, \dots X_k). \tag{5}$$

It will appear that the remainder ρ will involve essentially terms of the form

$$\sum_{k=1}^{n}\sum_{i=1}^{k-1}\|X_i\|_\infty\|E[X_k|\mathcal{F}_i]\|_\infty$$

(cf. the constant q in Theorem 1 and Theorem 7).

If the sequence is a random field this filtration will generally not be very useful because of the arbitrariness of the order on the variables, and we shall need to proceed in a such way that the sequence $\mathcal{F}_i$ in the interior sum actually depends on k. This is done as follows: to each index k we associate a reordering of the sequence which corresponds (hopefully) to increasing dependence with X_k; this brings on a new sequence, depending on k: X_j^k. This idea goes back to [6]. More precisely:

For any $1 \leqslant k \leqslant n$ is given a sequence X_j^k, $j = 1, \dots k$, which is a reordering of $(X_j,\ j = 1, \dots k)$ with $X_k^k = X_k$. We attach to each k a family of σ-algebras $(\mathcal{F}_j^k)_{j\leqslant k}$ such that

$$\mathcal{F}_j^k \supset \sigma(X_i^k, i \leqslant j), \quad j \leqslant k. \tag{6}$$

The σ-algebras $(\mathcal{F}_j)_{j\leqslant n}$ associated to the initial ordering are still defined by (5).

In particular, we have

$$\mathcal{F}_k^k \supset \mathcal{F}_k$$
$$\mathcal{F}_{k-1}^k \supset \mathcal{F}_{k-1}.$$

This is how we get a new constant $\tilde{q}$ (Eq. (26)) involving now terms of the form

$$\sum_{i=1}^{k-1}\|X_i^k\|_\infty\|E[X_k|\mathcal{F}_i^k]\|_\infty. \tag{7}$$

If (X_i) is a time series, it is natural to set $X_i^k = X_i$ and $\mathcal{F}_j^k = \mathcal{F}_j = \sigma(X_i, i \leqslant j)$: the superscripts can be dropped. Later on, the term "times series" will refer to to this situation, whereas the general case will be rather referred to as "random fields".

When dealing with mixing random fields of $\mathbb{R}^d$, each index k corresponds to some point P_k of the space where X_k sits; for each k, the sequence (X_j^k) will be typically obtained by sorting the original sequence $(X_j)_{j\leqslant k}$ in decreasing order of the distance $d(P_j, P_k)$. A simple example is the case of m-dependent fields indexed by $\mathbb{Z}^d$, that is, a process $X_a, a \in \mathbb{Z}^d$, such that the set of variables $X_A = \{X_a : a \in A\}$ is independent of X_B if the sets A and B have distance at least m; they are typically fields of the form $X_a = h(Y_{a+C})$

where Y_a is an i.i.d. random field, C a finite neighbourhood of 0 in $\mathbb{Z}^d$, and h a measurable function of $|C|$ variables.

We would like to point out that this framework covers quite different situations. For instance, in the Erdös-Rényi model of an unoriented random graph with n vertices, edges are represented by $\binom{n}{2}$ i.i.d. Bernoulli variables Y_{ab}, $1 \leqslant a < b \leqslant n$, with the convention $Y_{ab} = Y_{ba}$ and $Y_{aa} = 0$ (see [2] and references therein). The number of triangles (for instance) in such a model is

$$\sum_{a,b,c} Y_{ab}Y_{bc}Y_{ac} = \sum_{a,b,c} X_{abc}.$$

The process X is here an m-dependent process on the set of three element subsets of $\{1,\dots n\}$. We shall treat this example in Section 7.2.

Within this framework, we are able to control exponential moments of S with the help of formulas which generalize the Hoeffding and Bernstein inequalities for independent variables (Theorem 4 and Theorem 5). The bound of Theorem 4 involves $v = \sum_k \|E[X_k^2|\mathcal{F}_{k-1}]\|_\infty$, and a remainder term involving conditional expectations $\|E[X_k|\mathcal{F}_i^k]\|_\infty$. This is slightly unsatisfactory since it is known that the key quantity in the case of a martingale is the quadratic variation $\langle X\rangle = \sum_{k=1}^n E[X_k^2|\mathcal{F}_{k-1}]$, and in most cases effective bounds will actually involve $\|\langle X\rangle\|_\infty$, which is smaller than v. This is corrected in Theorem 1, where we give a result which generalizes what is known from martingale theory and improves on classical papers concerned with mixing [6]. However, inspection of the bounds shows that this improvement is really effective only if the conditional expectations $|E[X_k|\mathcal{F}_i^k]|$ are significantly smaller than $|X_k|$; if not, the only way to improve accuracy is to use the second order approach of Sections 5 and 6 briefly discussed below.

Second order results. By this terminology, we mean the following fact: the Hoeffding inequality (Equation 1 with $\rho = 0$) for instance is obtained from the exponential inequality

$$E[e^{tS}] \leqslant e^{t^2 \sum_i b_i^2/8}. \tag{8}$$

One obvious drawback of this upper bound is that when t tends to 0, it does not look like $1 + t^2 E[S^2]$. One would rather expect something like

$$E[e^{tS}] \leqslant e^{t^2 E[S^2]/2+Ct^3} \tag{9}$$

which has more interesting scaling properties; this approach would hopefully lead to significant improvements in a moderate deviation domain; this is what has been done in [5], but there S is an arbitrary function of independent variables, in the spirit of the McDiarmid inequality. In order to get closer, like in Equation (45) below, we have to pay with higher order extra terms: the remainder terms will not only contain conditional expectations $E[X_k|\mathcal{F}_i^k]$ but also conditional covariances; this will force us to consider for each pair of indices (i,j) another reordering of the sequence which corresponds to increasing dependence with the pair (X_i, X_j), and to introduce σ-fields $\mathcal{H}_k^{ij}$; we postpone details to Section 6.

In this context we shall give inequality of the form

$$|E[e^S] - e^{E[S^2]/2}| \leqslant e^{E[S^2]/2+Ct^3} \tag{10}$$

The paper is organized as follows. The three forthcoming sections deal with first order order exponential inequalities. A classical use of the exponential inequalities leads to Theorem 7 which generalizes the Bernstein and Hoeffding inequalities. An application to concentration inequalities and triangle counts is given in Section 4.

Section 5 to 7 are concerned with the second order approach, with applications to bounded difference inequalities and triangle counts.

In Section 8 we give some estimates under mixing assumptions.

2 First order approach for time series

This section is devoted to bounds for the Laplace transform of S in the case of time series (spatial processes are considered in the forthcoming section). The corresponding deviation probabilities will be obtained in Section 4.1 through classical arguments.

In Theorem 1 we present bounds which generalize known results concerning martingales.

In Theorem 4 we give a Hoeffding bound which is valid in both cases (time series and random fields), and Theorem 5 gives a Bennett bound for random fields which does not exactly generalizes (11) because the quadratic variation $\langle X\rangle$ is changed into the more drastic upper bound v.

The applicability of the following theorem depends on the way one can bound the quadratic variations involved. In the forthcoming examples, we shall consider only Equation (11) through a bound on $\|\langle X\rangle\|_\infty$; however Equations (12) and (13) have the advantage of not involving m.

Theorem 1. *Let us consider a sequence of centred random variables $(X_i)_{1\leqslant i\leqslant n}$ with the filtration defined by (5). We define*

$$
\begin{aligned}
&m = \sup_{1\leqslant k\leqslant n} \operatorname{ess\,sup} X_k\\
&S = \sum_{k=1}^n X_k\\
&[X] = \sum_{k=1}^n X_k^2\\
&\langle X\rangle = \sum_{k=1}^n E[X_k^2|\mathcal{F}_{k-1}]\\
&[X_+] = \sum_{k=1}^n (X_k)_+^2\\
&\langle X_-\rangle = \sum_{k=1}^n E[(X_k)_-^2|\mathcal{F}_{k-1}]\\
&q = \sum_{k=1}^n\sum_{i=1}^{k-1} \|X_i\|_\infty \|E[X_k|\mathcal{F}_i]\|_\infty + \tfrac{1}{3}\sum_{k=1}^n \|X_k\|_\infty \|E[X_k|\mathcal{F}_{k-1}]\|_\infty
\end{aligned}
$$

where the notation x_+^2 (resp. x_-^2) stands for $x^2 1_{x>0}$ (resp. $x^2 1_{x<0}$). Then

$$E\left[\exp\left(S - \frac{\langle X\rangle}{m^2}(e^m - m - 1)\right)\right] \leqslant e^{3q} \tag{11}$$

$$E\left[\exp\left(S - \frac{1}{2}[X_+] - \frac{1}{2}\langle X_-\rangle\right)\right] \leqslant e^{3q} \tag{12}$$

$$E\left[\exp\left(S - \frac{1}{6}[X] - \frac{1}{3}\langle X\rangle\right)\right] \leqslant e^{3q}. \tag{13}$$

If $S_k = X_1 + X_2 + \dots X_k$ is a supermartingale, these inequalities hold true with $q = 0$.

Remark. We recommend [1] for an account on recent work concerning exponential inequalities for martingales.

Proof. The key of the proof will be to use pairs of functions $\theta(x)$ and $\psi(x)$ such that

$$\psi \geqslant 0, \qquad e^{x-\theta(x)} \leqslant 1 + x + \psi(x). \tag{14}$$

These functions are meant to be $O(x^2)$ in the neighbourhood of 0. Three examples of such functions are

$$\theta(x) = 0, \quad \psi(x) = e^x - x - 1$$
$$\theta(x) = \zeta(x_+), \quad \psi(x) = \zeta(x_-), \quad \zeta(x) = e^{-x} + x - 1$$
$$\theta(x) = \frac{x^2}{6}, \quad \psi(x) = \frac{x^2}{3}.$$

Inequality (14) for these functions is proved in Proposition 15 of the Appendix. Set

$$T_k = \sum_{i=1}^{k} X_i - \theta(X_i) - \log(1 + \xi_i), \quad \text{where} \quad \xi_i = E[\psi(X_i)|\mathcal{F}_{i-1}].$$

Then

$$\begin{aligned} E[e^{T_n}] &= E[e^{X_n - \theta(X_n)}(1+\xi_n)^{-1} e^{T_{n-1}}] \\ &\leqslant E[(1 + X_n + \psi(X_n))(1+\xi_n)^{-1} e^{T_{n-1}}] \\ &= E[X_n(1+\xi_n)^{-1} e^{T_{n-1}}] + E[e^{T_{n-1}}] \end{aligned}$$

In the supermartingale case one gets

$$E[e^{T_n}] \leqslant E[e^{T_{n-1}}]. \tag{15}$$

In the general case we proceed as follows

$$\begin{aligned} E[e^{T_n}] &\leqslant E[X_n((1+\xi_n)^{-1} - 1)e^{T_{n-1}}] + E[X_n e^{T_{n-1}}] + E[e^{T_{n-1}}] \\ &= r_1 + r_2 + E[e^{T_{n-1}}]. \end{aligned} \tag{16}$$

We have

$$r_1 = E[E[X_n|\mathcal{F}_{n-1}] e^{T_{n-1}} \xi_n/(1+\xi_n)] \leqslant E[|E[X_n|\mathcal{F}_{n-1}]| e^{T_{n-1}}](\|\psi(X_n)\|_\infty \wedge 1). \tag{17}$$

The above defined function ψ is convex with $\psi(0) = 0$, $\psi(-1) \leqslant 1$ and $\psi(1) \leqslant 1$. Hence $|\psi(x)| \wedge 1 \leqslant |x|$ and therefore

$$r_1 \leqslant \|X_n\|_\infty \|E[X_n|\mathcal{F}_{n-1}]\|_\infty E[e^{T_{n-1}}].$$

Let $\Delta_i = T_i - T_{i-1}$; the second remainder is bounded as follows:

$$\begin{aligned} r_2 &= E \sum_{i=1}^{n-1} X_n(e^{T_i} - e^{T_{i-1}}) \\ &= E \sum_{i=1}^{n-1} X_n \tanh(\Delta_i/2)(e^{T_i} + e^{T_{i-1}}) \\ &\leqslant E \sum_{i=1}^{n-1} |E[X_n|\mathcal{F}_i] \tanh(\Delta_i/2)|(e^{T_i} + e^{T_{i-1}}) \\ &\leqslant \sum_{i=1}^{n-1} \|E[X_n|\mathcal{F}_i] \tanh(\Delta_i/2)\|_\infty 2 \sup_{j \leqslant n-1} E[e^{T_j}]. \end{aligned}$$

Equation (68) of the Appendix implies that $|\tanh(\Delta_i/2)| \leqslant 3\|X_i\|_\infty/2$, hence

$$r_2 \leqslant 3\Big(\sum_{i=1}^{n-1} \|X_i\|_\infty \|E[X_n|\mathcal{F}_i]\|_\infty\Big) \sup_{i \leqslant n-1} E[e^{T_i}]. \tag{18}$$

Finally, bringing together (16), (17) and (18)

$$
\begin{aligned}
E[e^{T_n}] &\leqslant (1+\rho_n) \sup_{i \leqslant n-1} E[e^{T_i}], \quad \rho_n = \|X_n\|_\infty \|E[X_n|\mathcal{F}_{n-1}]\|_\infty + 3\sum_{i=1}^{n-1} \|X_i\|_\infty \|E[X_n|\mathcal{F}_i]\|_\infty \\
&\leqslant \exp(\rho_n) \sup_{i \leqslant n-1} E[e^{T_i}]
\end{aligned}
$$

and we get by induction that

$$
\sup_{i \leqslant k} E[e^{T_i}] \leqslant \exp\Big(\sum_{i=1}^{k} \rho_i\Big).
$$

The right hand side can be set to 1 in the supermartingale case (cf. (15)). In particular

$$
E\Big[\exp\Big\{\sum_{i=1}^{n} X_i - \theta(X_i) - \log(1 + E[\psi(X_i)|\mathcal{F}_{i-1}])\Big\}\Big] \leqslant \exp(3q)
$$

hence

$$
E\Big[\exp\Big\{\sum_{i=1}^{n} X_i - \theta(X_i) - E[\psi(X_i)|\mathcal{F}_{i-1}]\Big\}\Big] \leqslant \exp(3q).
$$

This leads to the three bounds by using the three pairs of functions and by noticing that for $m \geqslant 0$ and $x \leqslant m$

$$
\varphi(x) \leqslant \varphi(m), \quad \varphi(x) = \frac{e^x - x - 1}{x^2} \tag{19}
$$

which is a consequence of L'Hospital's rule for monotonicity [20], and that for $x \geq 0$

$$
\zeta(x) \leqslant \frac{x^2}{2}
$$

since the function $x^2/2 - \zeta(x)$ has a non-negative derivative. □

Next theorem concerns an inequality which only involves $[X]$ (and not $\langle X \rangle$). The advantage is that no boundedness assumption is required since q does not appear. A definition is needed [1]:

Definition 2. *We shall say that an integrable random variable Y is heavy on left if*

$$
\forall a > 0, \quad E[T_a(Y)] \leqslant 0,
$$

where

$$
T_a(y) = \min(|y|, a)\,\mathrm{sign}(y)
$$

is the truncated version of y.

Many classical distributions satisfy this property for a reasonably large subset of parameter values [1]. Our definition differs slightly from [1] in the sense that we do not require Y to be centred; thus Theorem 3 below may be seen as an extension of Lemma C.1 of [1] to supermartingales:

Theorem 3. *Let us consider a sequence of non-necessarily centred random variables $(X_i)_{1 \leqslant i \leqslant n}$ with the filtration defined by (5). If for all k, the variable X_k is conditionally heavy on the left in the sense that*

$$
\forall a > 0, \quad E[T_a(X_k)|\mathcal{F}_{k-1}] \leq 0, \tag{20}
$$

then

$$
E\left[\exp\left(S - \tfrac{1}{2}[X]\right)\right] \leqslant 1. \tag{21}
$$

Proof. Using the classical inequality $\log\cosh x \le x^2/2$, and the identity $\cosh x = e^x/(1+\tanh x)$ we get

$$e^{x-x^2/2} \leqslant 1 + \tanh x.$$

Thus

$$\exp\Big(S - \tfrac{1}{2}[X]\Big) \leqslant \prod_{i=1}^{n}\Big(1 + \tanh X_k\Big).$$

On the other hand since $\tanh(x) = -\int_0^\infty \tanh''(a)T_a(x)da$, one has $E[\tanh(X_k)|\mathcal{F}_{k-1}] \le 0$, and the above product is clearly a supermartingale. □

Next theorem is a Hoeffding inequality. A weaker version of this theorem could be obtained by using (13):

Theorem 4. *Assume that we are in the setting described in the introduction, with a family of σ-fields satisfying (6). The variables X_k are centred. We define now $\tilde{q}$ as the first term in the expression of q:*

$$\tilde{q} = \sum_{k=1}^{n}\sum_{i=1}^{k-1} \|X_i\|_\infty \|E[X_k|\mathcal{F}_i]\|_\infty.$$

If the variables are lower and upper bounded with probability one:

$$a_i \leqslant X_i \leqslant a_i + b_i \tag{22}$$

the following inequality holds

$$E\Big[\exp\Big(S - \tfrac{1}{8}\sum_i b_i^2\Big)\Big] \leqslant e^{8\tilde{q}}. \tag{23}$$

In the supermartingale case (i.e. $E[X_i|\mathcal{F}_{i-1}] \leqslant 0$), this inequality remains true if we allow a_i and b_i to be an $\mathcal{F}_{i-1}$-measurable random variables.

Proof. We start, as in the proof of the Hoeffding inequality, with the following inequality based on the upper-bound of the exponential function by the chord over the curve on $[a, a+b]$:

$$e^x \leqslant \frac{(a+b)e^a - ae^{a+b}}{b} + x\frac{e^{a+b} - e^a}{b}, \qquad a \leqslant x \leqslant a+b.$$

It is well known that the first term of the right hand side, say e^c, is smaller than $\exp(b^2/8)$ independently of a (this a key step for proving the Hoeffding inequality, see for instance Appendix B of [21]). On the other hand, it is clear that $c \leqslant a + b$ (bound e^a with e^{a+b} in the expression of e^c). Hence, if we define c_i and d_i with the equations

$$c_i = \min\Big(\frac{b_i^2}{8}, a_i + b_i\Big),$$
$$d_i = e^{a_i}\frac{e^{b_i} - 1}{b_i}$$

we have

$$e^x \leqslant e^{c_i} + d_i x, \quad a_i \leqslant x \leqslant a_i + b_i.$$

Now let the random variables T_j be defined as

$$T_k = \sum_{i=1}^{k}(X_i - c_i)$$

We obtain

$$E[e^{T_n}] = E[e^{X_n} e^{T_{n-1}-c_n}] \leqslant E[(e^{c_n} + d_n X_n)e^{T_{n-1}-c_n}]$$

In the supermartingale case, the term involving d_n is $\leqslant 0$ and this equation gives immediately the result. We assume now that we are not necessarily in this case but a_i and b_i are deterministic. We can assume in addition, without loss of generality, that a_i and b_i are chosen so that (22) is tight. Notice that in this case we also have $a_i \leqslant 0 \leqslant a_i + b_i$ since $E[X_i] = 0$. The previous equation implies

$$\begin{aligned} E[e^{T_n}] &\leqslant d_n e^{-c_n} E[X_n e^{T_{n-1}}] + E[e^{T_{n-1}}] \\ &= d_n e^{-c_n} E \sum_{i=1}^{n-1} X_n (e^{T_i} - e^{T_{i-1}}) + E[e^{T_{n-1}}] \\ &= d_n e^{-c_n} r_2 + E[e^{T_{n-1}}]. \end{aligned} \tag{24}$$

Let $\Delta_i = T_i - T_{i-1} = X_i - c_i$; bounding r_2 as in the proof of Theorem 1 we get

$$\begin{aligned} r_2 &= E \sum_{i=1}^{n-1} X_n \tanh(\Delta_i/2)(e^{T_i} + e^{T_{i-1}}) \\ &\leqslant E \sum_{i=1}^{n-1} |E[X_n|\mathcal{F}_i]\Delta_i|(e^{T_i} + e^{T_{i-1}})/2 \\ &\leqslant \sum_{i=1}^{n-1} \|E[X_n|\mathcal{F}_i]\Delta_i\|_\infty \sup_{j \leqslant n-1} E[e^{T_j}] \end{aligned}$$

and since

$$|\Delta_i| \leqslant \max(a_i + b_i - c_i, c_i - a_i) \leqslant b_i \leqslant 2\|X_i\|_\infty$$

(b_i is the difference between the essential supremum and the essential infimum) we get that

$$r_2 \leqslant 2\rho_n \sup_{i \leqslant n-1} E[e^{T_i}], \quad \text{with} \quad \rho_n = \sum_{i=1}^{n-1} \|X_i\|_\infty \|E[X_n|\mathcal{F}_i]\|_\infty.$$

On the other hand, since $a_i \leqslant 0$, Equation (70) in Proposition 16 of the appendix leads to $d_n e^{-c_n} \leqslant 4$, and Equation (24) becomes finally

$$E[e^{T_n}] \leqslant (1 + 8\rho_n) \sup_{i \leqslant n-1} E[e^{T_i}] \leqslant e^{8\rho_n} \sup_{i \leqslant n-1} E[e^{T_i}].$$

Hence

$$E[e^{T_n}] \leqslant \exp\left(8 \sum_{i=1}^{n} \rho_i\right).$$

and we obtain (23) by using that $c_i \leqslant b_i^2/8$ in the expression of T_n. □

We introduce now a theorem which is almost a consequence of (11), except that q has been replaced by $\tilde{q}$. Its real interest is that unlike Theorem 1, this result will extend to the more general setting given by Equation (6).

Theorem 5. *Assume that we are in the setting described in the introduction, with a family of σ-fields satisfying (6). The variables X_k are centred. We define*

$$v = \sum_{k=1}^{n} \|E[X_k^2|\mathcal{F}_{k-1}]\|_\infty,$$

m as in Theorem 1, and $\tilde{q}$ as in Theorem 4. Then for any $t > 0$

$$E[e^S] \leqslant \exp\left(\frac{v}{m^2}(e^m - m - 1) + \tilde{q}\right). \tag{25}$$

Proof. We set

$$S_i = X_1 + X_2 + \cdots + X_i, \quad i \leqslant n$$

and $S_0 = 0$. Equation (19) implies that

$$e^{X_n} \leqslant 1 + X_n + X_n^2\varphi(m)$$

hence

$$\begin{aligned}
E[e^S] &\leqslant E[(1 + X_n + X_n^2\varphi(m))e^{S_{n-1}}] \\
&= E\sum_{i=1}^{n-1} X_n(e^{S_i} - e^{S_{i-1}}) + E[(1 + X_n^2\varphi(m))e^{S_{n-1}}] \\
&= E\sum_{i=1}^{n-1} X_n \tanh(X_i/2)(e^{S_i} + e^{S_{i-1}}) + \varphi(m)E[X_n^2 e^{S_{n-1}}] + E[e^{S_{n-1}}] \\
&\leqslant E\sum_{i=1}^{n-1} \|E[X_n|\mathcal{F}_i]X_i\|_\infty(e^{S_i} + e^{S_{i-1}})/2 + \varphi(m)E[E[X_n^2|\mathcal{F}_{n-1}]e^{S_{n-1}}] + E[e^{S_{n-1}}] \\
&\leqslant (1 + \tilde{q}^n + \varphi(m)v^n) \sup_{i\leqslant n-1} E[e^{S_i}] \\
&\leqslant e^{\tilde{q}^n + \varphi(m)v^n} \sup_{i\leqslant n-1} E[e^{S_i}]
\end{aligned}$$

where $\tilde{q}^n$ and v^n are the terms corresponding to $k = n$ in the definition of $\tilde{q}$ and v. This proves the result by induction. ☐

3 First order approach for spatial processes

The proofs Theorem 4 and Theorem 5 are based on a recursion like

$$E[e^{S_n}] \leqslant e^{q_n}E[e^{S_{n-1}}].$$

(Note that this is not the case in Theorem 1 since in its proof we consider $E[e^{T_n}]$ whose definition depends on the initial ordering of variables because of the conditionings.) At each step, the variables can be reordered in a more suitable order, as explained in the introduction. The proofs can be performed in the same way and we get

Theorem 6. *Assume that we are in the setting described in the introduction, with a family of σ-fields satisfying (6). The variables X_k are centred. We define*

$$\tilde{q} = \sum_{k=1}^{n}\sum_{i=1}^{k-1} \|X_i^k\|_\infty \|E[X_k|\mathcal{F}_i^k]\|_\infty \tag{26}$$

$$v = \sum_{k=1}^{n} \|E[X_k^2|\mathcal{F}_{k-1}]\|_\infty, \tag{27}$$

Then Theorem 4 and Theorem 5 hold true.

The m-dependent case. Let us consider the case of a m-dependent process in the following sense:

- for each k, there exist a set I_k of m_k indices (k included) such that

$$E[X_k|X_i, i \notin I_k] = 0. \tag{28}$$

One can start with an arbitrary initial ordering of the random variables. For any k, we can choose the sequence X_j^k in such a way that if $k - j \geqslant m_k$, $X_j^k = X_i$ for some $i \notin I_k$, in order to get

$$E[X_k|X_j^k, j \leqslant k - m_k] = 0. \tag{29}$$

$\mathcal{F}_i^k$ are the corresponding σ-fields. In this case

$$\tilde{q} = \sum_{k=1}^{n} \sum_{j<k} \|E[X_k|\mathcal{F}_j^k]\|_\infty \|X_j^k\|_\infty \leqslant \sup_j \|X_j\|_\infty \sum_{k=1}^{n} (m_k - 1)\|E[X_k|\mathcal{F}_{k-1}]\|_\infty$$

4 Applications of first order bounds

We postpone the discussion concerning spatial processes over a metric space to Section 8.

4.1 Deviation bounds

In this section we give the deviation inequalities that can be deduced from the preceding exponential inequalities. We generalize the Bernstein inequality in Equations (30) and (34), and the Hoeffding inequality in Equation (33); one could get Bennett inequalities through a similar process, we refer to Appendix B of [21]. In the martingale case, Equations (31) and (32) do not assume that the variables are bounded, but sums of squares are involved.

Theorem 7. *With the notations of Theorem 1 we have for any $A, y > 0$*

$$P(S \geqslant A, \langle X\rangle \leqslant y) \leqslant \exp\left(-\frac{A^2}{2(y+6q) + 2Am/3}\right) \tag{30}$$

$$P(S \geqslant A, [X_+] + \langle X_-\rangle \leqslant y) \leqslant \exp\left(-\frac{A^2}{2(y+6q)}\right) \tag{31}$$

$$P(S \geqslant A, [X] + 2\langle X\rangle \leqslant 3y) \leqslant \exp\left(-\frac{A^2}{2(y+6q)}\right). \tag{32}$$

With the notations of Theorem 4, 5 and 6 we have for any $A, y > 0$

$$P(S \geqslant A, \sum_i b_i^2 \leqslant 4y) \leqslant \exp\left(-\frac{A^2}{2y + 32\tilde{q}}\right) \tag{33}$$

$$P(S \geqslant A) \leqslant \exp\left(-\frac{A^2}{2(v + 2\tilde{q}) + 2Am/3}\right). \tag{34}$$

In the martingale case, (33) remains true if we allow a_i and b_i to be an $\mathcal{F}_{i-1}$-measurable random variable.

Remark. Equation (33) is analogous to Corollary 3(a) of [6].

Proof. Applying the bound (11) to the variables tX_i for some $t > 0$, we get

$$\begin{aligned}\log P(S \geqslant A, \langle X\rangle \leqslant y) &\leqslant \log E[\exp\{t(S - A) - \frac{t^2\langle X\rangle - t^2 y}{t^2 m^2}(e^{tm} - tm - 1)\}] \\ &\leqslant 3t^2 q + \frac{y}{m^2}(e^{tm} - tm - 1) - tA \\ &\leqslant \frac{y + 6q}{m^2}(e^{tm} - tm - 1) - tA.\end{aligned}$$

The optimization of this expression w.r.t. $t \geqslant 0$ is classical in the theory of Bennett and Bernstein inequalities and delivers (30); see for instance the Appendix B of [21]. The second inequality is deduced from (12) with the same method: for $V = [X_+] + \langle X_- \rangle$ or $V = ([X] + 2\langle X \rangle)/3$ one has

$$\log P(S \geqslant A, V \leqslant y) \leqslant \log E[e^{tS - tA - t^2(V-y)/2}] \leqslant 3t^2 q + y\frac{t^2}{2} - tA$$

and we take $t = A/(y + 8q)$.

Equations (33) and (34) are obtained similarly. $\square$

Deviation bounds for normalized sums. They are obtained through the following result (in the spirit of [7]):

Theorem 8. *If S and D are two random variables, $D \geqslant 0$, such that for any $t > 0$*

$$E\left[\exp\left(tS - \tfrac{1}{2}t^2 D\right)\right] \leqslant 1 \tag{35}$$

then, for any $p > 1$ and $x, y > 0$

$$P\Big(\frac{S}{\sqrt{a + D}} > x\Big) \leqslant e^{-\frac{x^2}{2p}}(xy)^{-1/p}, \qquad a = y^2 E[S_+^{1/(p-1)}]^{2(p-1)}. \tag{36}$$

Proof. Notice that for any variable $Y \sim \mathcal{N}(0, 1/a)$, $s \in \mathbb{R}$ and $d > 0$

$$E\left[e^{sY - \frac{1}{2}dY^2}\right] = \int \exp\Big(-\frac{d+a}{2}(y - \frac{s}{d+a})^2 + \frac{s^2}{2(d+a)}\Big)\frac{\sqrt{a}\,dy}{\sqrt{2\pi}} = \sqrt{\frac{a}{d+a}}\exp\Big(\frac{s^2}{2(d+a)}\Big).$$

Consequently, taking $(s, d) = (S, D)$ independent of Y in this equation and using (35) we get

$$E\Big[\sqrt{\frac{a}{D+a}}\exp\Big(\frac{S^2}{2(D+a)}\Big)1_{S>0}\Big] = E\Big[e^{YS - \frac{1}{2}Y^2 D}1_{S>0}\Big] \leqslant E\Big[e^{S|Y| - \frac{1}{2}DY^2}\Big] \leqslant 1.$$

Hence

$$\begin{aligned}
P\Big(\frac{S}{\sqrt{a+D}} > x\Big) &\leqslant E\Big[\Big(\frac{S}{\sqrt{D+a}}\Big)^{1/p}\exp\Big(\frac{S^2}{2p(D+a)}\Big)1_{S>0}\Big]x^{-1/p}e^{-\frac{x^2}{2p}} \\
&\leqslant E[S_+^{q/p}]^{1/q} E\Big[\sqrt{\frac{a}{D+a}}\exp\Big(\frac{S^2}{2(D+a)}\Big)1_{S>0}\Big]^{1/p} a^{-1/2p}x^{-1/p}e^{-\frac{x^2}{2p}} \\
&\leqslant E[S_+^{1/(p-1)}]^{(p-1)/p} a^{-1/2p} x^{-1/p} e^{-\frac{x^2}{2p}}.
\end{aligned}$$

We get (36) because of the specific choice of a. $\square$

This theorem can be applied with equations (12), (13) , (21) , (23) but not (11) or (25) because of the constant m in these equations which make the scaling impossible. For instance, (13) implies that for $t > 0$

$$E\left[\exp\left(tS - \frac{t^2}{6}[X] - \frac{t^2}{3}\langle X \rangle - 3t^2 q\right)\right] \leqslant 1$$

and (36) holds with

$$D = \tfrac{1}{6}[X] + \tfrac{1}{3}\langle X \rangle + 3q$$

where q is defined in Theorem 1.

4.2 Bounded difference inequalities

The above results lead straightforwardly to bounded difference inequalities by using a classical martingale argument of Maurey [19]. Equation (38) is the McDiarmid inequality [18]. Equation (39) is a Bernstein inequality in the same context; an advantage of this inequality is that it may give results even if the function D_k of (37) is not bounded

Theorem 9. *Let $Y = (Y_1, \dots Y_n)$ be a zero-mean sequence of independent variables with values in some measured space E. Let f be a measurable function on E^n with real values. We assume that for some functions $D_k(a,b)$ one has*

$$|f(y_1, \dots y_{k-1}, a, y_{k+1} \dots y_n) - f(y_1, \dots y_{k-1}, b, y_{k+1} \dots y_n)| \leqslant D_k(a,b). \tag{37}$$

Set

$$\begin{aligned} &S = f(Y) - E[f(Y)] \\ &m = \sup_k \|\Delta_k\|_\infty. \end{aligned}$$

Then for any $A > 0$

$$P(S \geqslant A) \leqslant \exp\left(-\frac{2A^2}{\sum_k \delta_k^2}\right), \qquad \delta_k = \|D_k\|_\infty \tag{38}$$

$$P(S \geqslant A) \leqslant \exp\left(-\frac{A^2}{2\sum_k {\delta'_k}^2 + 2Am/3}\right). \qquad \delta'_k = \|D_k(Y_k, Y'_k)\|_2 \tag{39}$$

where Y'_k is an independent copy of Y_k.

Remark. Let us mention that if f has the form $f(Y) = \sup_{g\in\Gamma} g(Y)$ for some finite class of functions Γ then, with obvious notations,

$$\begin{aligned} D_k(a,b) &= \sup_Y \Big| \sup_{g\in\Gamma} g(Y_1, \dots Y_{k-1}, a, Y_{k+1} \dots Y_n) - \sup_{g\in\Gamma} g(Y_1, \dots Y_{k-1}, b, Y_{k+1} \dots Y_n) \Big| \\ &\leqslant \sup_Y \Big| \sup_{g\in\Gamma} \Big\{ g(Y_1, \dots Y_{k-1}, a, Y_{k+1} \dots Y_n) - g(Y_1, \dots Y_{k-1}, b, Y_{k+1} \dots Y_n) \Big\} \Big| \\ &= \sup_{g\in\Gamma} D_k^g(a,b) \end{aligned}$$

in particular $\delta_k \leqslant \sup_{g\in\Gamma} \delta_k^g$. This is a classical argument in the theory of concentration inequalities.

Proof. We shall utilize (33) and (30) with

$$\begin{aligned} &X_k = E[f(Y)|\mathcal{F}_k] - E[f(Y)|\mathcal{F}_{k-1}] \\ &\mathcal{F}_k = \sigma(Y_1, \dots Y_k). \end{aligned}$$

We have already pointed out that $q = 0$ since X_k is a martingale difference. Let us define the random variables

$$\begin{aligned} &L_k = \inf_y E[f(Y_1, \dots Y_{k-1}, y, Y_{k+1} \dots Y_n)|\mathcal{F}_{k-1}] \\ &U_k = \sup_y E[f(Y_1, \dots Y_{k-1}, y, Y_{k+1} \dots Y_n)|\mathcal{F}_{k-1}]. \end{aligned}$$

The equation

$$L_k \leqslant E[f(Y)|\mathcal{F}_k] \leqslant U_k$$

implies

$$L_k - E[f(Y)|\mathcal{F}_{k-1}] \leqslant X_k \leqslant U_k - E[f(Y)|\mathcal{F}_{k-1}]$$

and since $U_k - L_k = \|D_k\|_\infty$ we can apply (33) with $b_k = \|D_k\|_\infty$ and get (38).

On the other hand, if we set

$$\Delta_k = f(Y) - f(Y_1, \dots Y_{k-1}, Y_k', Y_{k+1} \dots Y_n),$$

clearly X_k can be rewritten as

$$X_k = E[\Delta_k|\mathcal{F}_k]$$

hence $E[X_k^2 \; |\mathcal{F}_{k-1}] \leqslant E[\Delta_k^2 \; |\mathcal{F}_{k-1}] \leqslant {\delta_k'}^2$, and (39) follows from (30). □

Inequalities for suprema of U-statistics

For some problems of adaptive estimation and testing, it is very important to be able to control the supremum of U-statistics [12]. We give here a bound in this direction.

Consider a sequence of i.i.d. random variables $Y_1, \dots Y_n$ with values on some measurable space E and a finite family H of measurable symmetric functions on E^d and set for $h \in H$

$$\begin{aligned} &Y_I = \{Y_i, i \in I\}, \qquad I \subset \{1, \dots n\} \\ &Z_h(Y) = \frac{1}{\binom{n}{d}} \sum_{I \subset \{1,\dots n\}, |i|=d} h(Y_I) \\ &S = \sup_{h \in H} Z_h(Y) - E[\sup_{h \in H} Z_h(Y)] \end{aligned} \tag{40}$$

where the sum is restricted to the subsets with cardinality d; since the kernel h is symmetric there is no ambiguity regarding the notation $h(Y_A)$. We assume that h is centred:

$$E[h(Y_1, \dots Y_d)] = 0.$$

It is well known that if Z_h is non degenerate, that is

$$E[h(Y_1, \dots Y_d)|Y_1] \not\equiv 0,$$

then the variance of Z_h has order n^{-1}, cf [16] p.12. We give in the following corollary a deviation bound for S which corresponds to a Gaussian approximation with variance of the same order of magnitude. In the case of degenerate U-statistics we do not get good bounds; this is apparent in the case H has only one element since an abundant literature exists concerning deviation of degenerate U-statistics [14, 17, 4].

The function L below may be bounded by $2\|h\|_\infty$, and (41) should be generally good enough unless Y_1 takes a specific value with high probability in which case (42) may become significantly better:

Corollary 10. *If the symmetric function h satisfies for some function $L(x,y)$*

$$|h(y_1, \dots y_d) - h(y_1', y_2, \dots y_d)| \leqslant L(y_1, y_1').$$

Then, for any $A > 0$

$$P(S \geqslant A) \leqslant \exp\left(-\frac{2nA^2}{d^2\|L\|_\infty^2}\right), \tag{41}$$

$$P(S \geqslant A) \leqslant \exp\left(-\frac{nA^2}{d^2 E[L(Y_1, Y_1')^2] + 2Ad\|L\|_\infty/3}\right). \tag{42}$$

Proof. Fix k, a and b and set with $y^a = (y_1, \dots y_{k-1}, a, y_{k+1}, \dots y_n)$, then

$$\sup_{h,y} Z_h(y^a) - \sup_h Z_h(y^b) \leqslant \sup_h (Z_h(y^a) - Z_h(y^b))$$

thus, with the notation of Theorem 9,

$$|D_k(a,b)| \leqslant \sup_h |Z_h(y^a) - Z_h(y^b)|.$$

But

$$\begin{aligned} |Z_h(y^a) - Z_h(y^a)| &\leqslant \frac{1}{\binom{n}{d}} \Big| \sum_{I \ni k} h(y_I^a) - h(y_I^b) \Big| \\ &\leqslant \frac{1}{\binom{n}{d}} \binom{n-1}{d-1} L(a,b) \\ &= \frac{d}{n} L(a,b). \end{aligned}$$

Hence $D_k(a,b) \leqslant \frac{d}{n} L(a,b)$ and the result is now just a consequence of Theorem 9. □

5 Second order approach for time series

In this section we consider a sequence of centred random variables $(X_i)_{1 \leqslant i \leqslant n}$ with the filtration defined by (5). We set

$$S_k = \sum_{i=1}^{k} X_i. \tag{43}$$

The idea is to use the following decomposition

$$e^{S_n} = 1 - \sum_{j=1}^{n} e^{S_k} - e^{S_{k-1}} \simeq 1 - \sum_{k=1}^{n} X_k e^{S_k}$$

This is why we will need to control $X_k e^{S_k}$, for each k. This will be done by using again a similar decomposition:

Lemma 11. *Consider a sequence of centred random variables $(X_i)_{1 \leqslant i \leqslant n}$ with the filtration by (5) and the partial sums (43). Then*

$$\big| E[X_k e^{S_k}] - E[X_k S_k] E[e^{S_k}] \big| \leqslant w_k \sup_{j \leq k} E[e^{S_j}]$$

with

$$\begin{aligned} w_k = \tfrac{1}{2} \Big(\sum_{j \leqslant k} \|E[X_k|\mathcal{F}_j]\|_\infty \|X_j\|_\infty^2 \Big) &+ \Big(\sum_{j \leqslant k} \sum_{i \leqslant j} \|E[X_k X_j|\mathcal{F}_i] - E[X_k X_j]\|_\infty \|X_i\|_\infty \Big) \\ + \big(\sum_{j \leqslant k} |E[X_k S_{j-1}]| \|X_j\|_\infty \big). \end{aligned} \tag{44}$$

Proof.

$$\begin{aligned}
X_k e^{S_k} &= X_k + \sum_{j=1}^{k} X_k(e^{S_j} - e^{S_{j-1}} - X_j e^{S_j}) + \sum_{j=1}^{k} X_k X_j e^{S_j} \\
&= X_k + \sum_{j=1}^{k} X_k(e^{S_j} - e^{S_{j-1}} - X_j e^{S_j}) + \sum_{j=1}^{k}\sum_{i=1}^{j} X_k X_j (e^{S_i} - e^{S_{i-1}}) + \sum_{j=1}^{k} X_k X_j \\
&= X_k + \sum_{j=1}^{k} X_k(e^{S_j} - e^{S_{j-1}} - X_j e^{S_j}) + \sum_{j=1}^{k}\sum_{i=1}^{j} (X_k X_j - E[X_k X_j])(e^{S_i} - e^{S_{i-1}}) \\
&\quad - \sum_{j=1}^{k}\sum_{i=j+1}^{k} E[X_k X_j](e^{S_i} - e^{S_{i-1}}) + \sum_{j=1}^{k} E[X_k X_j](e^{S_k} - 1) + \sum_{j=1}^{k} X_k X_j \\
&= X_k + R_2 + R_3 + R_4 + E[X_k S_k]e^{S_k} - E[X_k S_k] + X_k S_k.
\end{aligned}$$

Concerning R_2, we notice that

$$|e^y - 1 - y| \leqslant \tfrac{1}{3} y^2 (\tfrac{1}{2} e^y + 1)$$

(this is due to the fact that for $\varepsilon = \pm 1$, $f_\varepsilon(y) = \frac{1}{3}y^2(\frac{1}{2}e^y + 1) + \varepsilon(e^y - 1 - y)$ is a convex function which satisfies $f'_\varepsilon(0) = f_\varepsilon(0) = 0$, thus $f_\varepsilon \geqslant 0$). Replacing y by $-y$ and multiplying by e^y we get

$$|1 - e^y + y e^y| \leqslant \tfrac{1}{3} y^2 (\tfrac{1}{2} + e^y)$$

hence, using this inequality in each term of R_2 with $y = X_j$:

$$|E[R_2]| \leqslant E \sum_{j\leqslant k} |E[X_k|\mathcal{F}_j]| \tfrac{1}{3} X_j^2 (\tfrac{1}{2} e^{S_{j-1}} + e^{S_j}) \leqslant \tfrac{1}{2} \Big(\sum_{j\leqslant k} \|E[X_k|\mathcal{F}_j]\|_\infty \|X_j\|_\infty^2 \Big) \sup_{j\leqslant k} E[e^{S_j}].$$

Using now the identity $e^u - 1 = (e^u + 1)\tanh(u/2)$, we have

$$\begin{aligned}
|E[R_3]| &= \Big| E \sum_{j\leqslant k}\sum_{i\leqslant j} (E[X_k X_j|\mathcal{F}_i] - E[X_k X_j]) \tanh(X_i/2)(e^{S_i} + e^{S_{i-1}}) | \\
&\leqslant \Big(\sum_{j\leqslant k}\sum_{i\leqslant j} \|E[X_k X_j|\mathcal{F}_i] - E[X_k X_j]\|_\infty \|X_i\|_\infty \Big) \sup_{j\leqslant k} E[e^{S_j}]
\end{aligned}$$

and similarly

$$\begin{aligned}
|E[R_4]| &= \Big| E \sum_{i=2}^{k} \Big(\sum_{j=1}^{i-1} E[X_k X_j] \Big) \tanh(X_i/2)(e^{S_{i-1}} + e^{S_i}) \Big| \\
&\leqslant \big(\sum_{i\leqslant k} |E[X_k S_{i-1}]| \|X_i\|_\infty \big) \sup_{j\leqslant k} E[e^{S_j}]
\end{aligned}$$

□

Theorem 12. *Consider a sequence of centred random variables $(X_i)_{1\leqslant i\leqslant n}$ with the filtration by (5) and the partial sums (43). Then*

$$\big| E[e^{S_n}] e^{-\frac{1}{2}E[S_n^2]} - 1 \big| \leqslant e^{\delta_0}(e^w - 1), \tag{45}$$

where

$$\delta_0 = \sum_{1 \leqslant j < k \leqslant n} E[X_k X_j]_- \qquad (x_- = -x 1_{x<0}) \tag{46}$$

$$w = \sum_{k=1}^{n} w_k \tag{47}$$

and w_k is given by (44). In particular, for any $t > 0$

$$E[e^{tS_n}] \leqslant e^{\frac{1}{2}t^2 E[S_n^2] + t^2\delta_0 + t^3 w}. \tag{48}$$

Remark. In the martingale case $\delta_0 = 0$ and

$$w_n = \tfrac{1}{2}\|X_n\|_\infty^3 + \sum_{i \leqslant n-1} \|E[X_n^2|\mathcal{F}_i] - E[X_n^2]\|_\infty \|X_i\|_\infty.$$

Proof. We consider the piecewise linear interpolation of the sequence S_n

$$\begin{aligned} S(t) &= S_k + (t-k)X_{k+1}, \qquad t \in [k, k+1] \\ &= S_{[t]} + (t - [t])X_{[t]+1}. \end{aligned}$$

Consider $t \in (n-1, n)$, i.e. $t = n - 1 + \tau$, $\tau \in (0,1)$, then

$$S(t) = S_{n-1+\tau} = S_{n-1} + \tau X_n$$

The previous lemma, applied where X_n is replaced with τX_n, and with $S(t)$ in place of S_n, implies that

$$\tau \big| E[S'(t)e^{S(t)}] - E[S'(t)S(t)]E[e^{S(t)}] \big| \leqslant \tau w(t) \sup_{s \leqslant t} E[e^{S(s)}]$$

with

$$w(t) = w_n.$$

We set

$$\begin{aligned} f(t) &= E[e^{S(t)}] \\ f^*(t) &= \sup_{s \leqslant t} f(s) \\ g(t) &= \tfrac{1}{2}E[S(t)^2]. \end{aligned}$$

then

$$\big| f'(t) - g'(t)f(t) \big| \leqslant w(t) f^*(t) \tag{49}$$

In particular

$$(fe^{-g})' = \tilde{w} f^* e^{-g}$$

for some function $\tilde{w}(t)$ such that

$$|\tilde{w}(t)| \leqslant w(t).$$

Hence

$$f(t)e^{-g(t)} = 1 + \int_0^t \tilde{w}(s) f^*(s) e^{-g(s)} ds. \tag{50}$$

We need now a first estimate of f^*. From (49) we get

$$f(t) \leqslant 1 + \int_0^t (g'(s)f(s) + w(s)f^*(s))ds$$

implying, if t_* is the index such that $f^*(t) = f(t_*)$, and if we set $g'(t)_+ = g'(t)1_{g'(t)>0}$,:

$$\begin{aligned} f^*(t) &\leqslant 1 + \int_0^{t_*} (g'(s)f(s) + w(s)f^*(s))ds \\ &\leqslant 1 + \int_0^t (g'(s)_+ + w(s))f^*(s)ds \end{aligned}$$

and by the Gronwall Lemma

$$f^*(t) \leq \exp\Big(\int_0^t (g'(s)_+ + w(s))ds\Big)$$

Inserting this back in (50):

$$\begin{aligned} |f(t)e^{-g(t)} - 1| &\leqslant \int_0^t w(s)\exp\Big(\int_0^s (g'(u)_+ + w(u))du\Big)e^{-g(s)}ds \\ &= \int_0^t w(s)\exp\Big(\int_0^s (|g'(u)|1_{g'(u)<0} + w(u))du\Big)ds \\ &\leqslant \exp\Big(\int_0^t |g'(u)|1_{g'(u)<0}du\Big)\int_0^t w(s)e^{\int_0^s w(u)du}ds \\ &\leqslant e^{\int_0^t |g'(u)|1_{g'(u)<0}du}\Big(e^{\int_0^t w(s)ds} - 1\Big). \end{aligned}$$

If $[u] = k-1$ then

$$g'(u) = E[X_k S_{k-1}] + (u-k+1)E[X_k^2] \geqslant -E[X_k S_{k-1}]_-$$

hence

$$\int_{k-1}^k |g'(u)|1_{g'(u)<0}du = \int_{k-1}^k (-g'(u))_+ du \leqslant E[X_k S_{k-1}]_-.$$

This implies (45). Equation (48) is obtained by noticing that (45) implies

$$E[e^{S_n}]e^{-\frac{1}{2}E[S_n^2]} - 1 \leqslant e^{\delta_0}e^w - 1$$

and by an elementary scaling argument. $\square$

6 Second order approach for spatial processes

As mentioned in the introduction, in order to get better results, we shall have to control conditional expectations of products $X_k X_j$; actually, our procedure will rather lead to products like $X_k X_j^k$; hence we are led to introduce for any pair (k, j) a sequence $X_i^{k,j}$ corresponding to increasing dependence with (X_k, X_j^k). More precisely:

Second order setting

(i) *For any* $1 \leqslant k \leqslant n$ *is given a sequence* X_j^k, $j = 1 \dots k$, *which is a reordering of* $(X_j,\ j = 1, \dots k)$ *with* $X_k^k = X_k$. *We attach to each* k *the* σ*-algebras*

$$\mathcal{F}_j^k \supset \sigma(X_i^k, i \leqslant j), \quad j \leqslant k. \tag{51}$$

(ii) *For* $1 \leqslant j \leqslant k \leqslant n$, *the sequence* $(X_i^{k,j})_{i \leqslant j}$ *is a reordering of* $(X_i^k)_{i \leqslant j}$ *and*

$$\mathcal{H}_i^{k,j} \supset \sigma(X_i^{k,j}, i \leqslant j). \tag{52}$$

In other words the σ-field $\mathcal{F}_j^k$ is made by taking off $\mathcal{F}_{j+1}^k$ the "closest" variable to X_k. The σ-field $\mathcal{H}_i^{k,j}$ is made by continuing this process after $\mathcal{F}_j^k$ (hence $i < j$) in a way which may depend on k and j.

This set-up is essentially, in a somewhat more general context, what is considered in [6]. For time series, we have $X_i^k = X_i$ and $\mathcal{F}_i^k = \mathcal{H}_i^{k,j} = \mathcal{F}_i = \sigma(X_l, l \leqslant i)$.

This setting is adequate for dealing with mixing random fields in which case each index k corresponds to some point P_k of the space; for each k, the sequence $(X_j^k)_j$ will be obtained by sorting the original sequence (X_j) in decreasing order of the distance $d(P_j, P_k)$, and $\mathcal{H}_j^k$ will be the σ-field generated by the random variables setting on the j more distant points from P_k, say $P_1^k, \dots P_j^k$; it is natural to define $X_i^{k,j}$ as X_l where P_l is the i-th more distant point from $\{P_k, P_j^k\}$, but the choice $X_i^{k,j} = X_i^k$, $\mathcal{H}_i^{k,j} = \mathcal{F}_i^k$, is typically suboptimal but good enough for random fields over the Euclidean space (see Section 8).

Theorem 13. *Consider a sequence of centred random variables* $(X_i)_{1 \leqslant i \leqslant n}$ *with the filtrations (51) and (52). Set*

$$\begin{aligned}
w &= \sum_{k=1}^{n} w_k \\
w_k &= \tfrac{1}{2} \sum_{j \leqslant k} \|E[X_k | \mathcal{F}_j^k]\|_\infty \|X_j^k\|_\infty^2 \ + \ \sum_{j \leqslant k} \sum_{i \leqslant j} \|E[X_k X_j^k | \mathcal{H}_i^{k,j}] - E[X_k X_j^k]\|_\infty \|X_i\|_\infty \\
&\quad + \sum_{j \leqslant k} |E[X_k S_{j-1}^k]| \|X_j^k\|_\infty \\
S_j^k &= \sum_{i=1}^{j} X_i^k.
\end{aligned}$$

Then (45) and (48) hold true.

Proof. We can work out Lemma 11 as before except that expressions like

$$\sum_{j=1}^{k} \sum_{i=1}^{j} X_k X_j (e^{S_i} - e^{S_{i-1}})$$

will become

$$\sum_{j=1}^{k} \sum_{i=1}^{j} X_k X_j^k (e^{S_i^{k,j}} - e^{S_{i-1}^{k,j}}), \quad S_i^{k,j} = \sum_{l=1}^{i} X_l^{k,j}.$$

and since $S_i^{k,j}$ is $\mathcal{H}_i^{k,j}$-measurable, these σ-fields appear in the expression of w_k. Since $S_k = S_k^{kk}$ and $S_j^k = S_j^{kj}$, we get as before

$$\left|E[X_k e^{S_k}] - E[X_k S_k]E[e^{S_k}]\right| \leqslant w_k \sup_{i\leqslant j\leqslant k} E\Big[e^{S_i^{kj}}\Big] \leqslant w_k \sup_{\alpha\in[0,1]^n} E\Big[e^{\sum_{j=1}^k \alpha_j X_j}\Big] \tag{53}$$

and the proof goes as the proof of Theorem 12 with a slight complication due to α. We still define

$$S(t) = \sum_{i\leqslant t} X_i + (t-[t])X_{[t]+1}$$
$$f(t) = E[e^{S(t)}]$$
$$g(t) = E[S(t)^2]$$

and, as we obtained (49) and (50), we get here

$$\left|f'(t) - g'(t)f(t)\right| \leqslant w(t)f^*(t) \tag{54}$$

which implies

$$f(t)e^{-g(t)} = 1 + \int_0^t \tilde{w}(s)f^*(s)e^{-g(s)}ds. \tag{55}$$

with $|\tilde{w}(s)| \leq w(s)$. In order to bound f^*, we define for any $\alpha \in \{0,1\}^n$

$$S_\alpha(t) = \sum_{i\leqslant t} \alpha_i X_i + (t-[t])\alpha_{[t]+1}X_{[t]+1}$$
$$f_\alpha(t) = E[e^{S_\alpha(t)}]$$
$$g_\alpha(t) = E[S_\alpha(t)^2]$$
$$f^*(t) = \sup_{\alpha\in[0,1]^n, s\leqslant t} f_\alpha(s). = \sup_{\alpha\in[0,1]^n} f_\alpha(t)$$

Equation (53) can be applied with $\alpha_k X_k$ instead of X_k and this leads to

$$|f'_\alpha(t) - g'_\alpha(t)f_\alpha(t)| \leqslant w_t f^*(t).$$

As before we get for any α

$$f_\alpha(t) \leqslant 1 + \int_0^t (g'_\alpha(s)f_\alpha(s) + w_s f^*(s))ds \leqslant 1 + \int_0^t ((g'_\alpha(s))_+ f_\alpha(s) + w_s f^*(s))ds$$

and this leads to

$$f^*(t) \leqslant 1 + \int_0^t (g'_*(s)f^*(s) + w_s f^*(s))ds$$

with now

$$g'_*(s) = \sup_\alpha g'_\alpha(s)_+.$$

The Gronwall Lemma leads to

$$f^*(t) \leq \exp\Big(\int_0^t (g'(s)_* + w(s))ds\Big).$$

We need now a bound on g_*':

$$
\begin{aligned}
g_*'(s) &= \sup_\alpha \alpha_{[s]} E[X_{[s]+1} S_\alpha(s)]_+ \\
&\leqslant \sum_{i \leqslant t} E[X_{[s]+1} X_i]_+ + (t-[t]) E[X_{[s]+1}^2] \\
&= g'(s) + \sum_{i \leqslant t} E[X_{[s]+1} X_i]_- .
\end{aligned}
$$

Inserting this back in (55):

$$
\begin{aligned}
|f(t)e^{-g(t)} - 1| &\leqslant \int_0^t w(s) \exp\Big(\int_0^s (g'(u)_* + w(u)) du\Big) e^{-g(s)} ds \\
&\leqslant e^{\int_0^t \sum_{i\leqslant t} E[X_{[s]+1}X_i]_- ds} \int_0^t w(s) e^{w(u)} duds \\
&= e^{\delta_0} \int_0^t w(s) e^{w(u)} duds.
\end{aligned}
$$

□

7 Applications of second order bounds

7.1 Deviation inequalities

In this section we give the deviation inequalities that can be deduced from the preceding exponential inequalities. The bound (56) is analogous to the bound or Corollary 3 (b) in [6] p.85, but in this paper the variance $E[S^2]$ is amplified with an extra factor $2e^2$ and w takes a slightly different value.

Theorem 14. *With the notations of Theorem 12 and Theorem 13, we have for any $A > 0$*

$$
P(S \geqslant A) \leqslant \exp\left(-\frac{A^2}{2E[S^2] + 4\delta_0 + 2\sqrt{2Aw/3}}\right). \tag{56}
$$

Remark. We recall that in the martingale case $\delta_0 = 0$ and

$$
w_n = \tfrac{1}{2}\|X_n\|_\infty^3 + \sum_{i\leqslant n} \|E[X_n^2|\mathcal{F}_i] - E[X_n^2]\|_\infty \|X_i\|_\infty.
$$

Proof. Using (48) we get

$$
\log P(S \geqslant A) \leqslant \log E[e^{t(S-A)}] \leqslant (\tfrac{1}{2}E[S^2] + \delta_0)t^2 + w\frac{t^3}{3} - tA.
$$

We choose $t = A/(E[S^2] + 2\delta_0 + \sqrt{2Aw/3})$; in particular $t \leqslant \sqrt{3A/2w}$ hence

$$
\log P(S \geqslant A) \leqslant (\tfrac{1}{2}E[S^2] + \delta_0)t^2 + \sqrt{2Aw/3}\,\frac{t^2}{2} - tA = -\frac{tA}{2} = -\frac{A^2}{2E[S^2] + 4\delta_0 + 2\sqrt{2Aw/3}}.
$$

□

7.2 m-dependency processes and triangle counts

7.2.1 m-dependent processes

Let us consider the case of a (m,n)-dependent process in the following sense:

- for each k, there exist a set I_k of m_k indices (k included) such that

$$E[X_k|X_i, i \notin I_k] = 0 \tag{57}$$

- for each pair (j,k), there exist a set $J_{j,k}$ of n_k indices such that

$$E[X_j X_k|X_i, i \notin J_{j,k}] = E[X_j X_k]. \tag{58}$$

One can start with an arbitrary initial ordering of the random variables. For any k, we can choose the sequence X_j^k in such a way that if $k-j \geqslant m_k$, $X_j^k = X_i$ for some $i \notin I_k$, in order to get

$$E[X_k|X_j^k, j \leqslant k - m_k] = 0. \tag{59}$$

Similarly we can choose the sequence $X_i^{k,j}$ in such a way that

$$E[X_j^k X_k|X_i^{k,j}, i \leqslant j - n_k] = E[X_j X_k]. \tag{60}$$

$\mathcal{F}_i^k$ and $\mathcal{H}_i^{k,j}$ are the corresponding σ-fields. In this case, for the calculation of w_k from Theorem 13 we notice that

$$\tfrac{1}{2}\sum_{j\leqslant k} \|E[X_k|\mathcal{F}_j^k]\|_\infty \|X_j^k\|_\infty^2 \leqslant \tfrac{1}{2} m_k \xi^2 \|X_k\|_\infty$$

and for the second term, if we set $\xi = \sup_i \|X_i\|_\infty$

$$\sum_{j\leqslant k}\sum_{i\leqslant j} \|E[X_k X_j^k|\mathcal{H}_i^{k,j}] - E[X_k X_j^k]\|_\infty \|X_i\|_\infty \leqslant n_k^2 \xi^2 \|X_k\|_\infty$$

and for the third one

$$\begin{aligned}
\sum_{j\leqslant k} |E[X_k S_j^k]| \|X_{j-1}^k\|_\infty &= \sum_{j=k-m_k+1}^{k} \sum_{i=k-m_k+1}^{j} |E[X_k X_i^k]| \|X_{j-1}^k\|_\infty \\
&\leqslant \xi^2 \|X_k\|_\infty \sum_{j=k-m_k+1}^{k} \sum_{i=k-n_k+1}^{j} 1 \\
&\leqslant \xi^2 \|X_k\|_\infty \tfrac{1}{2} m_k(m_k+1)
\end{aligned}$$

hence

$$w_k \leqslant \tfrac{1}{2}(m_k+1)^2 \xi^2 \|X_k\|_\infty + n_k^2 \xi^2 \|X_k\|_\infty.$$

Thus we can take

$$w = \xi^2 \sum_k \|X_k\|_\infty \left(\tfrac{1}{2}(m_k+1)^2 + n_k^2\right), \quad \xi = \sup_i \|X_i\|_\infty. \tag{61}$$

7.2.2 Triangle counts

We shall show that the standard Gaussian approximation is asymptotically valid for triangle counts in the moderate deviation domain.

In the Erdös-Rényi model of an unoriented random graph with n vertices, edges are represented by $\binom{n}{2}$ i.i.d. Bernoulli variables Y_{ab}, $1 \leqslant a < b \leqslant n$, with the convention $Y_{ab} = Y_{ba}$ and $Y_{aa} = 0$. The number of triangles in such a model is

$$Z = \sum_{\{a,b,c\}} Y_{ab} Y_{bc} Y_{ac}.$$

We set $p = E[Y_{12}]$ and

$$\begin{aligned} X_{abc} &= Y_{ab} Y_{bc} Y_{ac} - p^3 \\ S &= \sum_{\{a,b,c\}} Y_{ab} Y_{bc} Y_{ac} - p^3 = \sum_{\{a,b,c\}} X_{abc} = \sum_{\tau \in T} X_\tau \end{aligned} \tag{62}$$

where T is the set of subsets of $\{1, \dots n\}$ with three elements, $|T| = \binom{n}{3}$. For any $\tau = \{a, b, c\}$, define A_τ the set elements of T such that at least two points are in common with τ (τ, $\{a, b, d\}$, $\{a, d, c\}$, $\dots$); this makes $1 + 3(n-3)$ elements and $(X_\sigma)_{\sigma \notin A_\tau}$ is independent of X_τ. Hence we can take $m_k = 1 + 3(n-3)$.

Similarly, for any pair τ, τ', define by $B_{\tau,\tau'}$ the set of σ which have not two points in common with τ or τ'. Then $\sigma(X_\sigma, \sigma \in B_{\tau,\tau'})$ in independent of (X_τ, X'_τ). Since the complement of $B_{\tau,\tau'}$ has at most $2m_1$ variables, we can take $n_k = 2m_1$. Finally, since $\|X_i\|_\infty \leqslant 1$

$$w \leqslant \big(\tfrac{1}{2}(m_1 + 1)^2 + 4m_1^2\big)|T| \leqslant \big(\tfrac{9}{2}(3n)^2 \tfrac{1}{6} n^3 \leqslant 7n^5$$

and since $\delta_0 = 0$ (covariates are positively correlated) Equation (56) implies

$$P(S \geqslant A) \leqslant \exp\left(-\frac{A^2}{2Var(S) + 2\sqrt{14n^5 A/3}} \right). \tag{63}$$

Let us recall that (see [2])

$$Var(S) = \binom{n}{3}(p^3 - p^6) + \binom{n}{4}\binom{4}{2}(p^5 - p^6).$$

This has order n^4 due to the covariance terms. Let us briefly compare with the bound of [2]; this paper delivers a bound for $P(S \geqslant A)$ which is slightly larger than

$$\exp\left(-\frac{A^2}{6nE[Z] + 16\sqrt{E[Z]}A} \right), \qquad E[Z] = \frac{n(n-1)(n-2)}{6} p^3 \tag{64}$$

(the actual formula is much more complicated; we have used that $\min(a^{-1}, b^{-1}) \leqslant 2/(a+b)$ to obtain this from Theorem 18 of [2]). One has $6nE[Z] \geqslant \frac{2}{p^2 - p^3} 2Var(S)$. For p fixed and n large, the square root term in (63) is residual if $A \ll n^3$; this is the moderate deviation case since the centring term in (62) has order n^3 (notice that $S \leqslant n^3$ w.p.1), and we get the right variance. In (64), a change occurs when $n^{5/2} \ll A$, and if we set $A = Bn^{5/2}$ with B large, (64) leads to $\exp(-cnB)$ while (62) behaves like $\exp(-cnB^2)$.

8 Evaluation of constants under φ-mixing assumptions

We give here informally some arguments to convince the reader that under standard φ-mixing assumptions the constant q has the same order as the variance of the sum, and that w will be small.

For more details about mixing we refer to [6], [10] and [3] (particularly Section 8 concerning random fields).

The φ-mixing constant between two σ-fields $\mathcal{A}$ and $\mathcal{B}$ is defined as

$$\varphi(\mathcal{A},\mathcal{B}) = \sup_{A\in\mathcal{A},B\in\mathcal{B}} |P(B|A) - P(B)|.$$

It is well known, see reference [22] p.27 or [15] p.278, that this implies that if Z is a zero-mean $\mathcal{B}$-measurable random variable

$$\|E[Z|\mathcal{A}]\|_\infty \leqslant 2\varphi(\mathcal{A},\mathcal{B})\|Z\|_\infty.$$

Assume that X is a field over a part of $\mathbb{Z}^d$: each variable X_k of the field sits on some $P_k \in \mathbb{Z}^d$. $\mathcal{F}_j^k$ is the σ-field generated by the j more distant points from P_k different from X_1,...X_{k-1}, and we take for simplicity $\mathcal{H}_i^{k,j} = \mathcal{F}_i^k$.

Denote by P_j^k the the jth closest point to P_k, when the points $P_1,\dots P_{k-1}$ have been excluded. Thus X_j^k sits on P_{k-j+1}^k. The distance of P_k to P_j^k, is at least $cj^{1/d}$, for some constant c. This implies that standard φ-mixing assumptions between $\sigma(X_k)$ and $\mathcal{F}_j^k$ can be rewritten as

$$\varphi(\mathcal{F}_j^k,\sigma(X_k)) \leqslant \varphi_{\infty,1}((k-j)^{1/d}) \tag{65}$$

for some decreasing function $\varphi_{\infty,1}$ (for example exponential decay holds for finite range[1] shift-invariant Gibbs random fields [13] pp. 158-159; this contains a lot of examples). The subscripts ∞ and 1 on φ mean that there is no restriction on the number of random variables contained in the first σ-field, $\mathcal{F}_j^k$, and there is only 1 variable in the second, $\sigma(X_k)$; we use this traditional notation, in particular for compatibility with [6].

On the other hand for $i < j$, $j - i$ is smaller than the number of points in the annulus $\{x : \|P_j^k - P_k\| \leqslant \|x - P_k\| \leqslant \|P_i^k - P_k\|\}$, in particular, for some c

$$j - i \leqslant c\|P_i^k - P_k\|^{d-1}\|P_j^k - P_i^k\|.$$

Hence $\|P_j^k - P_i^k\|$ is at least $c(j-i)i^{-1+1/d}$ for some c. This implies that standard φ-mixing assumptions between $\sigma(X_k, X_j^k)$ and $\mathcal{F}_i^k$ can be rewritten as

$$\varphi(\mathcal{F}_i^k,\sigma(X_k,X_j^k)) \leqslant \varphi_{\infty,2}((j-i)(k-i+1)^{-1+1/d}), \quad i \leqslant j \tag{66}$$

for some decreasing function $\varphi_{\infty,2}$.

Equations (65) and (66) will imply, for $i \leqslant j \leqslant k$, and any measurable bounded functions f and g

$$\begin{aligned}
&\|E[f(X_k)|\mathcal{F}_j^k]\|_\infty \leqslant 2\|f(X_k)\|_\infty\varphi_{\infty,1}((k-j)^{1/d})\\
&\|E[g(X_k,X_j^k)|\mathcal{F}_i^k] - E[g(X_k,X_j^k)]\|_\infty \leqslant 2\|g(X_k,X_j^k)\|_\infty\varphi_{\infty,2}((j-i)(k-i+1)^{-1+1/d}).
\end{aligned}$$

The first equation leads to

$$\begin{aligned}
&\|E[X_k|\mathcal{F}_j^k]\|_\infty \leqslant 2\|X_k\|_\infty\varphi_{\infty,1}((k-j)^{1/d})\\
&\|E[X_kX_j^k|\mathcal{F}_i^k] - E[X_kX_j^k]\|_\infty \leqslant 2m\|E[X_k|\mathcal{F}_j^k]\|_\infty \leqslant 4m\|X_k\|_\infty\varphi_{\infty,1}((k-j)^{1/d})
\end{aligned}$$

with $m = \sup_i \|X_i\|_\infty$, and the second

$$\|E[X_kX_j^k|\mathcal{F}_i^k] - E[X_kX_j^k]\|_\infty \leqslant 2m\|X_k\|_\infty\varphi_{\infty,2}((j-i)(k-i+1)^{-1+1/d}).$$

[1] This means the existence of a constant c such that if $k,l \in \mathbb{Z}^d$ are such that $d(k,l) > c$ then X_k and X_l are independent conditionally to $(X_j)_{j\notin\{k,l\}}$.

Hence

$$\|E[X_kX_j^k|\mathcal{F}_i^k]-E[X_kX_j^k]\|_\infty \leqslant 4m\|X_k\|_\infty \min(\varphi_{\infty,2}((j-i)(k-j+1)^{-1+1/d}),\varphi_{\infty,1}((k-j)^{1/d})).$$

We get with $m_1=\sum\|X_i\|_\infty$, and setting $\varphi(x)=\varphi([x])$

$$\begin{aligned}\tilde{q} &\leqslant 2\sum_k\sum_{i<k}\|X_i^k\|_\infty\|E[X_k|\mathcal{F}_i^k]\|_\infty\\ &\leqslant 2m\sum_k\sum_{i<k}\|X_k\|_\infty\|\varphi_{\infty,1}(k-i)\\ &\leqslant 2m\sum_k\|X_k\|_\infty\int_0^\infty\varphi_{\infty,1}(x^{1/d})dx\\ &=2dm_1m\int_0^\infty\varphi_{\infty,1}(x)x^{d-1}dx.\end{aligned}$$

The integral is essentially the quantity $B(\phi)$ of [6], and Equations (33) and (34) may be seen as improvements over (b)(i) and (ii) of Corollary 4 of [6]. The constant w in Theorem 13 is the sum of three terms. The first one is

$$\begin{aligned}\sum_k\sum_{j=1}^k\|E[X_k|\mathcal{F}_j^k\|_\infty\|E[X_j^k\|_\infty^2 &= \tfrac{1}{2}m_3+\tfrac{1}{2}m^2\sum_{j<k}\|E[X_k|\mathcal{F}_j^k\|_\infty,\quad m_3=\sum\|X_i\|_\infty^3\\ &\leqslant \tfrac{1}{2}m_3+m^2\sum_{j<k}\|X_k\|_\infty\varphi_{\infty,1}((k-j)^{1/d})\\ &\leqslant \tfrac{1}{2}m_3+m^2\sum_k\sum_{j>0}\|X_k\|_\infty\varphi_{\infty,1}(j^{1/d})\\ &\leqslant \tfrac{1}{2}m_3+m^2m_1\sum_{j>0}\int_{j-1}^j\varphi_{\infty,1}(x^{1/d})dx\\ &= \tfrac{1}{2}m_3+dm_1m^2\int_0^\infty y^{d-1}\varphi_{\infty,1}(y)dy\end{aligned}$$

The second term contributing to w is

$$\begin{aligned}&\sum_{i\leqslant j\leqslant k}\|E[X_kX_j^k|\mathcal{F}_i^k]-E[X_kX_j^k]\|_\infty\|X_i\|_\infty\\ &\quad\leqslant 4m^2\sum_{i\leqslant j\leqslant k}\|X_i\|_\infty\min(\varphi_{\infty,2}((j-i)(k-j+1)^{-1+1/d}),\varphi_{\infty,1}((k-j)^{1/d}))\\ &\quad= 4m^2\sum_{i\leqslant k}\sum_{0\leqslant j\leqslant k-i}\|X_i\|_\infty\min(\varphi_{\infty,2}((k-i-j)(j+1)^{-1+1/d}),\varphi_{\infty,1}(j^{1/d}))\\ &\quad= 4m^2\sum_{i,k\geqslant 0}\sum_{0\leqslant j\leqslant k}\|X_i\|_\infty\min(\varphi_{\infty,2}((k-j)(j+1)^{-1+1/d}),\varphi_{\infty,1}(j^{1/d}))\\ &\quad\leqslant 4m^2m_1\sum_{k\geqslant 0}\sum_{0\leqslant j\leqslant k}\min(\varphi_{\infty,2}((k-j)(j+1)^{-1+1/d}),\varphi_{\infty,1}(j^{1/d}))\\ &\quad\leqslant 4m^2m_1\sum_{0\leqslant j\leqslant (k-1)/2}\varphi_{\infty,2}((k-j)(j+1)^{-1+1/d})+4m^2m_1\sum_{0\leqslant (k-1)/2<j\leqslant k}\varphi_{\infty,1}(j^{1/d})\\ &\quad= 4m^2m_1T_1+4m^2m_1T_2\end{aligned}$$

For the first term, since $j+1 \leqslant k-j$ and $\frac{1}{2}(k+1) \leqslant k-j$:

$$\begin{aligned}
T_1 &\leqslant \sum_{0 \leqslant j \leqslant (k-1)/2} \varphi_{\infty,2}((k-j)^{1/d}) \\
&\leqslant \sum_{k \geqslant 1} \tfrac{1}{2}(k+1)\varphi_{\infty,2}((\tfrac{1}{2}(k+1))^{1/d}) \\
&\leqslant \sum_{k \geqslant 1} \tfrac{1}{2} \int_k^{k+1} (z+1)\varphi_{\infty,2}((\tfrac{1}{2}z)^{1/d}) dz \\
&= \int_{1/2}^{\infty} (2u+1)\varphi_{\infty,2}(u^{1/d}) du \\
&= d \int_{1/2}^{\infty} (2y^d+1) y^{d-1} \varphi_{\infty,2}(y) dy
\end{aligned}$$

and for the second one

$$\begin{aligned}
T_2 &= \sum_{0 \leqslant (k-1)/2 < j \leqslant k} \varphi_{\infty,1}(j^{1/d}) \\
&= \sum_{j \geqslant 1} (j+1)\varphi_{\infty,1}(j^{1/d}) \\
&\leqslant \sum_{j \geqslant 1} \int_{j-1}^{j} (u+2)\varphi_{\infty,1}(u^{1/d}) du \\
&= d \int_0^{\infty} (y^d+2) y^{d-1} \varphi_{\infty,1}(y) dy
\end{aligned}$$

The third term in w is bounded as

$$\begin{aligned}
\sum_k \sum_{j \leqslant k} |E[X_k S_{j-1}^k]| \|X_j^k\|_\infty &\leqslant m^2 \sum_k \sum_{j=1}^{k} \sum_{i<j} \|E[X_k|\mathcal{F}_i^k]\|_\infty \\
&= m^2 \sum_k \sum_{i<k} (k-i) \|E[X_k|\mathcal{F}_i^k]\|_\infty \\
&\leqslant 2m^2 \sum_k \sum_{i<k} (k-i) \|X_k\|_\infty \varphi_{\infty,1}((k-i)^{1/d}) \\
&\leqslant 2m^2 \sum_k \sum_{j>0} j \|X_k\|_\infty \varphi_{\infty,1}(j^{1/d}) \\
&\leqslant 2m^2 m_1 \sum_{j>0} \int_{j-1}^{j} (x+1)\varphi_{\infty,1}(x^{1/d}) dx \\
&= 2dm_1 m^2 \int_0^{\infty} (1+y^d) y^{d-1} \varphi_{\infty,1}(y) dy.
\end{aligned}$$

Putting everything together we finally get

$$w \leqslant \tfrac{1}{2} m_3 + 11 d m^2 m_1 \int_0^{\infty} (1+y^d) y^{d-1} (\varphi_{\infty,1}(y) + \varphi_{\infty,2}(y)) dy$$

The integral in the right hand side is essentially the $D(\phi)$ of [6] page 86. If we refer to the independent case (X_k of order $1/\sqrt{n}$) the factor $m^2 m_1$ is of order $1/\sqrt{n}$, what makes the factor of t^3 in (48) residual as far as t is smaller than $\sqrt{n}$ (moderate deviations).

A Technical inequalities

Proposition 15. *The three pairs of functions*

$$\theta(x) = 0, \quad \psi(x) = e^x - x - 1$$
$$\theta(x) = \zeta(x_+), \quad \psi(x) = \zeta(x_-), \quad \zeta(x) = e^{-x} + x - 1$$
$$\theta(x) = \frac{x^2}{6}, \quad \psi(x) = \frac{x^2}{3}$$

satisfy

$$e^{x-\theta(x)} \leqslant 1 + x + \psi(x). \tag{67}$$

We have

$$\left| \tanh\left(\tfrac{1}{2}(x - \theta(x) - \log(1 + \psi(y)))\right)\right| \leqslant \tfrac{3}{2}m, \quad |x|, |y| \leqslant m. \tag{68}$$

Proof. Everything will be more or less based on the inequality $e^x \leqslant 1 + x$.

Equation (67) is obvious for the first pair of functions. In the second case we have only to check for $x > 0$; since in this case $x - \theta(x) = 1 - e^{-x}$ this reduces to proving that:

$$1 - e^{-x} \leqslant \log(1 + x).$$

The function $\log(1 + x) + e^{-x} - 1$ has a derivative $(1 + x)^{-1} - e^{-x}$ which is $\geqslant 0$ since $e^x \geqslant 1 + x$; hence the inequality is satisfied. The third case is the non negativity of the function

$$f(x) = 1 + x + \frac{x^2}{3} - e^{x-x^2/6}.$$

This function satisfies

$$f'(x) = 1 + \frac{2x}{3} - \Big(1 - \frac{x}{3}\Big)e^{x-x^2/6}$$
$$f''(x) = \frac{2}{3} - \Big(\Big(1 - \frac{x}{3}\Big)^2 - \frac{1}{3}\Big)e^{x-x^2/6} = \frac{2}{3}\Big(1 - (1 - x + x^2/6)e^{x-x^2/6}\Big)$$

which is non negative since $1 - (1 - u)e^u \geqslant 0$; f is convex. Since $f'(0) = 0$ and $f(0) = 0$, we conclude that f is non negative.

For the last inequality, we start with an upper bound on $\psi(y)$. In the first case

$$\psi(y) \leqslant \psi(|y|) \leqslant \psi(m) \leqslant e^m - 1.$$

In the second case $\psi(y) \leqslant \psi(-m) \leqslant e^m - 1$ and in the third case $\psi(y) \leqslant m^2/3 \leqslant e^m - 1$ (because of the expansion of the exponential); in any case we have

$$\log(1 + \psi(y)) \leqslant m. \tag{69}$$

On the one hand

$$\tanh(\tfrac{1}{2}(x - \theta(x) - \log(1 + \psi(y)))) \leqslant \tanh(\tfrac{1}{2}x) \leqslant m/2$$

and on the other hand, thanks to (69), using that $\theta(x) \leqslant x^2/2$

$$\begin{aligned}\tanh(\tfrac{1}{2}(\log(1 + \psi(y)) - x + \theta(x))) &\leqslant \tanh(m + m^2/4)\\ &\leqslant \min(m + m^2/4, 1)\\ &\leqslant \tfrac{5}{4}m\end{aligned}$$

by considering separately the cases $m < 1$ and $m > 1$.

□

Proposition 16. *For any $a \leqslant 0 \leqslant b$ one has*

$$e^a \, \frac{e^b - 1}{b} \leqslant 4 \exp\left(\min\left(\frac{b^2}{8}, a+b\right)\right). \tag{70}$$

Proof. We have indeed

$$\begin{aligned}
\exp\left(-\frac{b^2}{8}\right) \frac{e^b - 1}{b} &= \exp\left(-\frac{b^2}{8}\right) \frac{(e^{b/2}+1)(e^{b/2}-1)}{b} \\
&= e^{-b^2/8}(e^{b/2}+1)^2 \frac{\tanh(b/4)}{b} \\
&\leqslant (e^{b/2 - b^2/16} + e^{-b^2/16})^2 \, \frac{1}{4} \\
&\leqslant \frac{1}{4}(e+1)^2 \\
&< 4.
\end{aligned}$$

Hence

$$e^a \frac{e^b - 1}{b} \leqslant 4 \exp\left(\frac{b^2}{8}\right).$$

And clearly

$$e^a \frac{e^b - 1}{b} = e^{a+b} \frac{1 - e^{-b}}{b} \leqslant e^{a+b}.$$

□

References

[1] B. Bercu, A. Touati, Exponential inequalities for self-normalized martingales with applications, *Ann. Appl. Probab.* 13 (2008), no. 5, 1848–1869. MR2462551

[2] S. Boucheron, G. Lugosi, P. Massart, Concentration inequalities using the entropy method, *Ann. Probab.* 31 (2003), no. 3, 1583–1614. MR1989444

[3] R.C. Bradley, Basic properties of strong mixing conditions. A survey and some open questions. Update of, and a supplement to, the 1986 original. MR0899990 *Probab. Surv.* 2 (2005), 107–144.

[4] J. Bretagnolle, A new large deviation inequality for U-statistics of order 2, *ESAIM: Probability and Statistics* 3 (1999), p. 151-162. MR1742613

[5] O. Catoni, Laplace transform estimates and deviation inequalities. *Ann. Inst. H. Poincaré Probab. Statist.* 39 (2003), no. 1, 1–26. MR1959840

[6] J. Dedecker, Exponential Inequalities and Functional central limit theorems for random fields, *ESAIM* 5 (2001), 77-104. MR1875665

[7] V.H. De la Peña, M.J. Klass, and T.L. Lai, Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws, *Ann. Probab.*, Vol. 32, pp. 1902–1933, 2004.

[8] V.H. De la Peña, M.J. Klass, and T.L. Lai, Pseudo-maximization and self-normalized processes. *Probab. Surv.* 4 (2007), 172–192.

[9] B. Delyon, *Exponential inequalities for sums of weakly dependent variables*, *Electron. J. Probab.* 14 (2009), no. 28, 752–779. MR2495559

[10] P. Doukhan, *Mixing: Properties and Examples*, *Lecture Notes in Stat.* 85, 1994. MR1312160

[11]

[12] M. Fromont, B. Laurent, Adaptive goodness-of-fit tests in a density model. *Ann. Statist.* 34 (2006), no. 2, 680–720. MR2281881

[13] H-O. Georgii, *Gibbs Measures and Phase Transitions*, de Gruyter Studies in Mathematics 9, Walter de Gruyter & Co., 1988. MR0956646

[14] E. Giné, R. Latala, J. Zinn, Exponential and moment inequalities for U-statistics. *High dimensional probability, II*, 13–38, Progr. Probab., 47 (2000), Birkhauser Boston, Boston, MA, 2000. MR1857312

[15] P. Hall, C.C. Heyde, *Martingale Limit Theory and its Applications*, Academic Press, 1980. MR0624435

[16] A.J. Lee, *U-Statistics : theory and practice*, Statistics: Textbooks and Monographs, 110. M Marcel Dekker, 1990. MR1075417

[17] P. Major, A multivariate version of Hoeffding's inequality, *E.C.P.* 11 (2006), 220—229. MR2266713

[18] C. McDiarmid, On the method of bounded differences, *Surveys in Combinatorics 1989*, 148–188, London Math. Soc. Lecture Note Ser., 141, Cambridge Univ. Press. 1989. MR1036755

[19] B. Maurey, Construction de suites symétriques. *C. R. Acad. Sci.*, Paris, Sér. A 288 (1979), 679-681. MR0533901

[20] I. Pinelis, l'Hospital type rules for monotonicity, with applications, *J. Inequal. Pure Appl. Math.* 3 (2001), No. 1, Article 5. `http://www.emis.de/journals/JIPAM/article158.html`

[21] D. Pollard , *Convergence of Stochastic Processes*, Springer, 1984. MR0762984

[22] E. Rio, *Théorie asymptotique des processus aléatoires faiblement dépendants*, Springer, 2000. MR2117923